Dipl.-Inf. Nils Höller

Übungen zur Vorlesung

Mobile und Verteilte Datenbanken

WS 2009/2010

Übung 12 – Anfrageverarbeitung in Wireless Sensor Networks

Aufgabe 1: Ausfallwahrscheinlichkeit TAGk

In der Vorlesung wurde eine Erweiterung des TAG Prinzips vorgestellt, in das Vorhandensein mehrerer Nachbarn ausgenutzt wird, um Aggregations-Teilergebnisse zu verteilen. Das Ziel dieser Verteilung ist, vorhandene Redundanz auszunutzen, um fehlertolerante Aggregation zu ermöglichen.

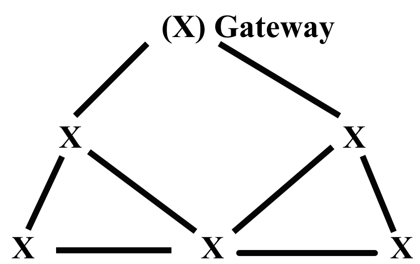
In dieser Aufgabe sollen Sie die Verbesserung der Fehlertoleranz abhängig von der Hinzunahme zusätzlicher Elternknoten abschätzen.

Es gelten folgende Notationen für die Verfahren:

- TAG: Aggregation über einen Vaterknoten
- TAG(k): Aggregation durch Verteilung auf k Vaterknoten

Als Aggregationfunktion soll die Funktion COUNT betrachtet werden.

Gegeben sei folgendes Sensor Netzwerk mit den eingezeichneten bidirektionalen Links:



a) Führen Sie schrittweise am Beispiel die Aggregation für TAG und TAG2 aus

Gegeben seien folgende Parameter für das Sensor Netzwerk:

- Eine Nachricht wird mit der Wahrscheinlichkeit p versendet
- Nachrichtenverluste sind unabhängige Ereignisse

Betrachten Sie nun folgende Fragestellungen:

b) Was ist der Erwartungswert für den Count Wert den Knoten s zu einem einzelnen Vater sendet?

c) Wie lässt sich die Varianz abschätzen?

- Tipp: Die Varianz beschreibt die Streuung (hier: um den eigentlichen Count Wert), d.h. die Abweichung vom Erwartungswert
- Tipp2: Es liegt eine Bernoulli Verteilung vor, d.h. wir haben ein Experiment in dem es zwei Ereignisse geben kann: 1.) Es kommt das Teilergebnis an oder 2.) Es kommt kein Teilergebnis an. In Fall 2) wird dann ein Teilergebnis von 0 angenommen.
- Tipp3: Die Varianz für eine Zufallsvariable X berechnet sich nach
  $VAR(X) = E((X-E(X))^2) = E(X^2) – (E(X))^2$
- Tipp4: Die Varianz für die Standard Bernoulli Verteilung beträgt $VAR(X) = pq$

d) Wie verändert sich der Erwartungswert, wenn Knoten s das Teilergebnis c/2 (also jeweils den halben Count Wert) zu 2 Vätern sendet? Hinweis: Es liegt eine Linearität für den Erwartungswert vor, d.h. der Gesamterwartungswert für das Endergebnis ist die Summe der Erwartungswerte für beide Väter.

e) Wie lässt sich die neue Gesamtvarianz abschätzen? Hinweis: Auch hier gilt die Linearität.

f) Was lässt sich nun schlussfolgern?

Abschließende Bemerkung: Die Annahme, dass Nachrichtenverluste unabhängige Ereignisse sind, ist in der Realität nicht korrekt. Sie wurde zur Vereinfachung gewählt. In einem realen Szenario verstärken sich obige Ergebnisse.

Aufgabe 2: Synopsis Diffusion

In der Vorlesung wurde das approximative Aggregationsverfahren Synopsis Diffusion vorgestellt.

a) Erläutern sie noch mal zum Verständnis, was man unter ODI Korrektheit versteht und wie man einen ODI Beweis führt.

b) Uniform Sample:
Gesucht ist ein Synopsis zur Berechnung des Uniform Sample der Länge k, d.h. einer gleichverteilten Probe der Länge k, welche demnach die Werte im Sensornetzwerk bestmöglich widerspiegelt. Beschreiben Sie die Synopsis mit den einzelnen Synopsisfunktionen und führen Sie den ODI Beweis
Tipp: Denken Sie daran, dass das Ausnutzen der Funktion MAX ODI Correctness impliziert.

Dipl.-Inf. Nils Höller

Übungen zur Vorlesung

Mobile und Verteilte Datenbanken

Lösung

WS 2009/2010

Übung 12 – Anfrageverarbeitung in Wireless Sensor Networks

Lösung1:

1. **TAG: Minimal spannenden Baum definieren und Werte zur Wurzel hochzählen Gesamtergebnis = 6**
2. **E(X) = c * p (P(0) = (1-p), P(COUNT) = c)**
3. **VAR(X) = c^2*p*(1-p) = E(X^2) – E(X)^2 = c*p – (c*p)^2 = c^2 * p * (1-p)**
4. **E(X) = c/2*p + c/2*p = c*p**
5. **VAR(X) = 2*(c/2)^2 * p * (1-p) = c^2/2 * p * (1-p)**
6. **Die Varianz mit mehreren Vätern ist viel geringer als mit einem Vater bei gleich bleibendem Gesamtwert. Dieses Ergebnis wird in realen Szenarien noch verstärkt, da die Unabhängigkeitsannahme nicht mehr gilt.**

Lösung2:

Synopsis Diffusion (Definitionen siehe Script)

Uniform Sample:

- Each sensor has a reading. Compute a uniform sample of a given size k.
  a) Synopsis: a sample of k tuples.

  SG(): output (value, r, id), where r is a uniform random number in range [0, 1].

  SF(): output the k tuples with the k largest r values. If there are less than k tuples in total, out them all.

  SE(): output the values in s.
- ODI-correctness is implied by “MAX” and union operation in SF().
- Correctness: the largest k random numbers is a uniform k sample.